

## Consensus-based method for risk adjustment for surgery for congenital heart disease

Kathy J. Jenkins, MD, MPH<sup>a</sup>  
Kimberlee Gauvreau, ScD<sup>a</sup>  
Jane W. Newburger, MD, MPH<sup>a</sup>  
Thomas L. Spray, MD<sup>b</sup>  
James H. Moller, MD<sup>c</sup>  
Lisa I. Iezzoni, MD, MS<sup>d</sup>

See related editorial on page 6.

**Objective:** The aim was to develop a consensus-based method of risk adjustment for in-hospital mortality among children younger than 18 years after surgery for congenital heart disease (designated RACHS-1).

**Methods:** An 11-member national panel of pediatric cardiologists and cardiac surgeons used clinical judgment to place surgical procedures into six risk categories. Categories were refined after review of information from the Pediatric Cardiac Care Consortium and three statewide hospital discharge data sets. The effects of including additional clinical variables were explored by comparing areas under receiver-operator characteristic curves.

**Results:** Among 4602 surgical patients in the Pediatric Cardiac Care Consortium data set and 4493 in the hospital discharge data, 3767 (81.9%) and 3832 (85.3%), respectively, had a single cardiac procedure, and 98.5% and 89.2%, respectively, were able to be assigned to one of six risk categories defined by the panel. Mortality rates showed expected trends ( $P < .001$ ). For the Pediatric Cardiac Care Consortium data, mortality rates were 0.4% in category 1, 3.8% in 2, 8.5% in 3, 19.4% in 4, and 47.7% in 6; rates were similar in the hospital discharge data. There were too few cases in category 5 to estimate mortality rates. In multivariable models, younger age, prematurity, and the presence of a major noncardiac structural anomaly added to the risk of in-hospital death predicted by risk category alone. Best performance was obtained when cases with multiple procedures were placed in the risk category of the most complex procedure.

**Conclusion:** The RACHS-1 method should adjust for baseline risk differences and allow meaningful comparisons of in-hospital mortality for groups of children undergoing surgery for congenital heart disease.

From the Department of Cardiology, Children's Hospital, Boston,<sup>a</sup> the Department of Cardiovascular Surgery, Children's Hospital of Philadelphia,<sup>b</sup> the Department of Pediatrics, Minnesota Children's Hospital, Minneapolis, Minn,<sup>c</sup> and the Division of General Medicine, Beth Israel Deaconess Medical Center, Boston.<sup>d</sup>

Supported by National Institutes of Health, National Heart, Lung, and Blood Institute, grant K08HL2936-01 (K.J.J.) and by the Kobren Fund (K.G.).

Received for publication March 15, 2001; revisions requested May 10, 2001; revisions received June 18, 2001; accepted for publication July 25, 2001.

Address for reprints: Kathy J. Jenkins, MD, MPH, Department of Cardiology, Children's Hospital, 300 Longwood Ave, Boston, MA 02115 (E-mail: jenkins@cardio.tch.harvard.edu).

J Thorac Cardiovasc Surg 2002;123:110-8

Copyright © 2002 by The American Association for Thoracic Surgery.

0022-5223/2002 \$35.00 + 0 12/1/119064

doi:10.1067/mtc.2002.119064

Nearly every stakeholder in the pediatric health care environment—patient, parent, practitioner, program director, payor, and policy maker—needs to understand the outcomes of therapy for serious disease in childhood. Because each child is different, however, any evaluation of treatments must take into account important differences that increase the risk for poor outcomes among some children but not others. Each congenital heart defect is a rare condition. In combination, such defects comprise a leading cause of death in childhood. Despite an increasing attention

to outcomes research in this field,<sup>1,2</sup> no method of risk adjustment for congenital heart disease is currently accepted. In previous work we applied a technique of grouping procedures together that showed promise according to internal validation<sup>3</sup> but had little consensus among practitioners. Others have used a similar approach.<sup>4,5</sup> Additional attempts to adjust for risk have been limited to specific age groups or procedures.<sup>6,7</sup>

Acknowledging both the difficulty and the need, we have attempted to create, on the basis of consensus and clinical judgment, a formal, iterative method of risk adjustment for short-term mortality from all forms of surgery for congenital heart disease. The extreme diversity of the conditions to be included and the limited availability of large data sets precluded the use of simple empirical analysis. We took advantage of the facts that a single procedure is used for the correction of most defects (eg, tetralogy of Fallot repair for a tetralogy of Fallot defect) and also that many anatomic abnormalities are palliated by similar strategies (eg, Fontan palliation for multiple forms of single ventricle). This allowed a reduction in the amount of diversity among conditions by using procedure type as a surrogate for anatomy. A parallel exists between the creation of a risk-adjustment method and the clinician's evaluation of potential surgical risks that is part of the day-to-day practice of pediatric cardiology and cardiac surgery; this parallel strengthened our choice of a judgment-based approach. The resulting method, which was created with sustained input from a panel of national experts from 10 institutions, was evaluated with two large multi-institutional data sets. The panel then used selective judgment-based refinements to correct deficiencies in the initial performance of the risk-adjustment method. The overall result is a valid clinical research tool, designated the Risk Adjustment for Congenital Heart Surgery 1 (RACHS-1), that may be widely applicable to the evaluation of differences in outcomes in both existing and future data sets.

## Methods

### Panel of Experts

The RACHS-1 method was created with the clinical expertise of a nationally representative 11-member panel of experts. The volunteer members of our panel included both pediatric cardiologists and pediatric cardiac surgeons (Table 1).

### Consensus-Based Process

The risk-adjustment method was created at three face-to-face meetings of the panel (May 12, 1997, November 10, 1997, and November 9, 1998) and with both paper and electronic surveys. Throughout the process panel members chose to focus on creating a tool that would (1) be widely applicable, (2) use data that are relatively easy to obtain, and (3) allow meaningful comparisons among groups of patients with congenital heart disease. Because risk-adjustment methods are by definition outcome-specific,<sup>8</sup> the panel first selected the particular outcome of interest for the pro-

**TABLE 1. Panel of experts**

Eugene H. Blackstone, MD
Departments of Thoracic and Cardiovascular Surgery
Cleveland Clinic Foundation, Cleveland, Ohio
David E. Fixler, MD
Division of Cardiology
University of Texas Southwest Medical Center, Dallas, Tex
Howard P. Gutgesell, MD
Department of Pediatric Cardiology
University of Virginia Medical Center, Charlottesville, Va
Frank L. Hanley, MD
Division of Cardiothoracic Surgery
University of California San Francisco Medical Center,
San Francisco, Calif
Thomas J. Kulik, MD
Department of Pediatric Cardiology
C.S. Mott Children's Hospital, Ann Arbor, Mich
John E. Mayer, Jr, MD
Department of Cardiovascular Surgery
Children's Hospital, Boston, Mass
James H. Moller, MD
Division of Cardiology
University of Minneapolis Hospital, Minneapolis, Minn
Jane W. Newburger, MD, MPH
Department of Cardiology
Children's Hospital, Boston, Mass
Jan M. Quaegebeur, MD
Department of Cardiothoracic Surgery
Columbia University Medical Center, New York, NY
Thomas L. Spray, MD
Department of Cardiothoracic Surgery
Children's Hospital of Philadelphia, Philadelphia, Pa
Roberta G. Williams, MD
Department of Pediatrics
University of North Carolina at Chapel Hill School of Medicine,
Chapel Hill, NC

ject. After discussion of various alternatives at the initial meeting (eg, functional status, neurologic outcome, late mortality), the panel chose to focus on short-term mortality for surgery for congenital heart disease in children younger than 18 years. Recognizing that information is available and generally reliable for in-hospital deaths in all US hospitals, the panel chose to create a method that would specifically adjust for differences in in-hospital mortality rates among groups of children undergoing surgery for congenital heart disease.

Panel members then discussed clinical factors that made children undergoing surgery for a congenital heart defect more likely to die before discharge. Early in the process they concluded that the most important risk factor for groups of children was the type of surgical procedure performed, rather than the specific anatomic diagnosis or other clinical variables. To deal with the large number of diverse procedures, panel members chose to group procedures that carried similar risks of in-hospital mortality into risk categories.

Each panel member was provided with a list of 207 surgical procedures drawn from both *Current Procedural Terminology 4* and International Classification of Diseases, Ninth Revision, Clinical Modification (*ICD-9-CM*) codes. Members were asked to group procedures with similar risks of in-hospital death. The total number of risk categories to be used was not specified. Because panel members differed in the number of risk categories created (from 4 to 6), a theoretic 0 to 1 scale was divided into subintervals to represent the number of groups chosen by a particular member. Individual procedure "risks" were then assigned as the midpoint of the appropriate subinterval. These "normalized" risk values were averaged across panel members, and procedures were ordered from lowest to highest risk with these averages. Theoretic cut points for risk categories were then determined on the basis of totals of 3 to 6 categories, according to a linear 0 to 1 scale.

The ranked surgical procedures, individual panel member groupings, and theoretic risk categories were then distributed to the full panel at a second face-to-face meeting. Members examined the results and decided by consensus to create six distinct risk categories. The panel then deliberated about those procedures with considerable intrapanel variability regarding the appropriate risk grouping, as well as procedures at or near the cut points for category change. By consensus some procedures were moved to a more appropriate category or were deemed "uncategorizable."

After the second meeting, custom software was distributed that allowed each panel member to review the categorization of each procedure and to vote to change the categorization if he or she believed it to be incorrect. As part of this process members were allowed to stratify procedures into two or more categories according to other variables, such as age or cardiac diagnosis. Members were also permitted to add procedures for which no *Current Procedural Terminology 4* or *ICD-9-CM* code existed. Changes considered to be appropriate by a majority of the panel were incorporated. As a result of this process, five procedures were stratified by age at surgery and two were stratified by specific cardiac diagnosis. Members were then asked to specify additional risk factors that should be included in the risk-adjustment method.

At a third meeting, the panel reviewed the overall behavior of the preliminary risk categories using data from two large multi-institutional data sets, as described here. Mortality rates, 95% confidence intervals, and numbers of cases performed in each data set were examined for each procedure. After reviewing this information, the panel revised the categorizations of some procedures because the actual mortality rate differed considerably from the initial judgment about risk for death. For example, the panel had initially placed systemic to pulmonary artery shunt in category 2 but chose to move it to category 3 after reviewing the data. After this process only a few procedures remained in risk category 5. This was noted by the panel and discussed. Because the risk of death for these procedures was still considered distinct from those of adjacent groups, the panel chose to retain the separate category; however, the panel recognized that the ability to make inferences for this group would necessarily be limited. Members then discussed which additional risk factors to include in the final risk-adjustment method and how to account for "combination procedures" in which two or more distinct cardiac surgical procedures were performed simultaneously. The panel had some uncertainty about these decisions and therefore chose to rely

on an empirical analysis of information from the two data sets to make these determinations.

### Data Sources

Data for the refinement and validation part of this project were obtained from two different sources. One data set was obtained from the Pediatric Cardiac Care Consortium (PCCC),<sup>9</sup> which collects and analyzes data for all cardiac procedures performed at member institutions on an annual basis for the purpose of quality improvement. Information for cardiac surgical procedures performed in children younger than 18 years at 32 institutions in calendar year 1996, without institutional or patient identifier information, was obtained and analyzed specifically for this project. Permission to use data from the PCCC was obtained from its board of directors. The high quality of the PCCC data set is ensured in several ways: the program coordinator manually reviews all cardiac diagnosis and procedure codes for consistency among institutions; information is entered into the computer system by two different individuals, and the two versions of the data are compared electronically; computer programs check for out of range values, nonsequential dates, and missing data; and independent verification of the number of procedures performed at each institution in a given year is obtained. A second data set consisting of hospital discharge data purchased from three states (Illinois 1994, Massachusetts 1995, and California 1995) included a similar number of cases as the PCCC 1996 information. The discharge data consisted of specific information for each hospital discharge occurring within each state for the specified year in standard formats and had not been analyzed previously. Hospital discharge data sets are collected by designated state agencies and are available to the public, with permission, for research purposes; individual agencies are responsible for the quality and completeness of all information. No approval is required at our institution for analysis of data available in the public domain.

### Case Selection

For both the PCCC and hospital discharge data sets, all patients younger than 18 years with operative or procedure codes indicating surgical repair of a congenital heart defect were selected. Patients undergoing cardiac transplantation (PCCC operative codes 1952, 19521-15222, and 1953; *ICD-9-CM* procedure code 37.5) and neonates 30 days of age or younger with patent ductus arteriosus as an isolated cardiac defect (PCCC operative codes 1761, 17611-17612, 1762, and 17621-17622; *ICD-9-CM* diagnosis code 747.0 and procedure code 38.85) were eliminated. Patients with patent ductus arteriosus and birth weight less than 2500 g had previously been excluded from the PCCC data set. Within the hospital discharge data set, transcatheter interventions were eliminated by excluding cases in which the sole cardiac procedure was atrial septal defect closure (*ICD-9-CM* diagnosis code 745.5 and procedure code 35.51 or 35.71), ventricular septal defect closure (*ICD-9-CM* diagnosis code 745.4 and procedure code 35.53 or 35.72), patent ductus arteriosus closure (*ICD-9-CM* diagnosis code 747.0 and procedure code 38.85), atrial septectomy (*ICD-9-CM* procedure code 35.42), or vessel repair or occlusion (*ICD-9-CM* procedure code 38.85 or 39.59) with concomitant

**TABLE 2. Individual procedures by risk category**

Risk category 1
Atrial septal defect surgery (including atrial septal defect secundum, sinus venosus atrial septal defect, patent foramen ovale closure)
Aortopexy
Patent ductus arteriosus surgery at age >30 d
Coarctation repair at age >30 d
Partially anomalous pulmonary venous connection surgery
Risk category 2
Aortic valvotomy or valvuloplasty at age >30 d
Subaortic stenosis resection
Pulmonary valvotomy or valvuloplasty
Pulmonary valve replacement
Right ventricular infundibulectomy
Pulmonary outflow tract augmentation
Repair of coronary artery fistula
Atrial septal defect and ventricular septal defect repair
Atrial septal defect primum repair
Ventricular septal defect repair
Ventricular septal defect closure and pulmonary valvotomy or infundibular resection
Ventricular septal defect closure and pulmonary artery band removal
Repair of unspecified septal defect
Total repair of tetralogy of Fallot
Repair of total anomalous pulmonary veins at age >30 d
Glenn shunt
Vascular ring surgery
Repair of aorta-pulmonary window
Coarctation repair at age ≤30 d
Repair of pulmonary artery stenosis
Transection of pulmonary artery
Common atrium closure
Left ventricular to right atrial shunt repair
Risk category 3
Aortic valve replacement
Ross procedure
Left ventricular outflow tract patch
Ventriculomyotomy
Aortoplasty
Mitral valvotomy or valvuloplasty
Mitral valve replacement
Valvectomy of tricuspid valve
Tricuspid valvotomy or valvuloplasty
Tricuspid valve replacement
Tricuspid valve repositioning for Ebstein anomaly at age >30 d
Repair of anomalous coronary artery without intrapulmonary tunnel
Repair of anomalous coronary artery with intrapulmonary tunnel (Takeuchi)

catheterization codes (*ICD-9-CM* procedure codes 37.21-37.23, 88.42-88.44, or 88.50-88.58) and no code for cardiopulmonary bypass (*ICD-9-CM* procedure code 39.61). Further details of this process are available on request.

**TABLE 2. Cont'd**

Closure of semilunar valve, aortic or pulmonary
Right ventricular to pulmonary artery conduit
Left ventricular to pulmonary artery conduit
Repair of double-outlet right ventricle with or without repair of right ventricular obstruction
Fontan procedure
Repair of transitional or complete atrioventricular canal with or without valve replacement
Pulmonary artery banding
Repair of tetralogy of Fallot with pulmonary atresia
Repair of cor triatriatum
Systemic to pulmonary artery shunt
Atrial switch operation
Arterial switch operation
Reimplantation of anomalous pulmonary artery
Annuloplasty
Repair of coarctation and ventricular septal defect closure
Excision of intracardiac tumor
Risk category 4
Aortic valvotomy or valvuloplasty at age ≤30 d
Konno procedure
Repair of complex anomaly (single ventricle) by ventricular septal defect enlargement
Repair of total anomalous pulmonary veins at age ≤30 d
Atrial septectomy
Repair of transposition, ventricular septal defect, and subpulmonary stenosis (Rastelli)
Atrial switch operation with ventricular septal defect closure
Atrial switch operation with repair of subpulmonary stenosis
Arterial switch operation with pulmonary artery band removal
Arterial switch operation with ventricular septal defect closure
Arterial switch operation with repair of subpulmonary stenosis
Repair of truncus arteriosus
Repair of hypoplastic or interrupted arch without ventricular septal defect closure
Repair of hypoplastic or interrupted aortic arch with ventricular septal defect closure
Transverse arch graft
Unifocalization for tetralogy of Fallot and pulmonary atresia
Double switch
Risk category 5
Tricuspid valve repositioning for neonatal Ebstein anomaly at age ≤30 d
Repair of truncus arteriosus and interrupted arch
Risk category 6
Stage 1 repair of hypoplastic left heart syndrome (Norwood operation)
Stage 1 repair of nonhypoplastic left heart syndrome conditions
Damus-Kaye-Stansel procedure

### Statistical Methods

Data from the PCCC and hospital discharge data sets were analyzed separately. Newly created algorithms were used to assign cases to the six risk categories. Patients listed as undergoing com-

**TABLE 3. Mortality rates by risk category**

Risk category	Single procedure		Combination procedure		Total mortality rate (%)
	No. of cases	Mortality rate (%)	No. of cases	Mortality rate (%)	
PCCC data					
1	961	0.4	3	0.0	0.4
2	1222	3.8	213	3.8	3.8
3	1205	8.5	318	13.5	9.5
4	191	19.4	85	18.8	19.2
5	2	—	2	—	—
6	128	47.7	40	45.0	47.0
Hospital discharge data					
1	922	0.3	0	—	0.3
2	1290	3.3	64	3.1	3.3
3	1038	6.8	128	3.9	6.5
4	128	16.4	35	31.4	19.6
5	0	—	0	—	—
6	41	41.5	0	—	41.5

binations of surgical procedures were initially assigned to multiple risk categories, one for each individual procedure. Patients undergoing a procedure that was deemed uncategorizable by the panel of experts or that had not been assigned to a risk category were eliminated from further analysis. The details of the case assignment process are also available on request.

Analyses were initially restricted to patients undergoing a single cardiac procedure. Mortality rates and 95% confidence intervals were calculated for each risk category. Differences in mortality across categories were evaluated with a  $\chi^2$  test for trend.

Risk categories 2, 3, 4, 5, and 6 were then used as binary covariates in a logistic regression model predicting mortality, with category 1 as the reference group. Odds ratios measuring the risk of death for each category relative to that in category 1 were estimated. Risk category was then retained in the model, and additional clinical factors were introduced one at a time to explore whether they improved the model's ability to predict in-hospital mortality. The likelihood ratio test was used to assess whether the inclusion of each clinical factor in addition to risk category was more predictive of in-hospital mortality than was risk category alone. The receiver-operator characteristic (ROC) curve corresponding to each model was also examined. The area under the ROC curve (also known as the *c statistic*) quantifies how well the model is able to discriminate between patients who did and did not die; an area of 0.5 indicates a model with no predictive power, whereas a model that perfectly predicts outcome every time has area 1.0. When included in the multivariable model with risk category, the single factor that provided the most additional predictive information about in-hospital mortality was retained in the model; the remaining factors were then reevaluated. This process was repeated until none of the remaining factors contributed significantly to the prediction of mortality.

Patients undergoing combinations of surgical procedures were assigned separately to a risk category for each distinct procedure. To evaluate how best to incorporate these cases, we examined various levels of adjustment. Combination procedures were first

placed in the risk category of the single most complex procedure. By comparing the areas under ROC curves, we evaluated whether increasing the highest risk category by 1 or summing the individual risk categories changed the ability of the model to predict mortality beyond the original categorization. The presence of combination procedures was also considered as a binary covariate in the multivariable models.

The final model for each data set was evaluated for goodness of fit with the Hosmer-Lemeshow test, which measures the predictive ability of a model by comparing the observed and expected numbers of deaths for cases in 10 groups with similar predicted probabilities of in-hospital mortality, from those with the lowest predicted probabilities of death to those with the highest. Within each of these deciles the observed mortality rate was then plotted against the expected mortality rate based on the multivariable model. A line was fitted through these points by means of linear regression weighted by the number of cases in each decile, and the coefficient of determination ( $R^2$ ) was calculated.

## Results

### Risk Categories

The individual procedures included in each of the final six risk categories are listed in Table 2. Most procedures were in categories 1 through 3. Extremely few procedures were placed in category 5. As previously mentioned, the panel considered moving these procedures to either of the adjacent categories but on discussion determined that the risk of in-hospital death for these procedures was different from those in either category 4 or category 6; they therefore chose to retain the separate categorization.

### Data Analysis

Within the PCCC data set, a total of 4602 patients who underwent surgical repair of a congenital heart defect were found.

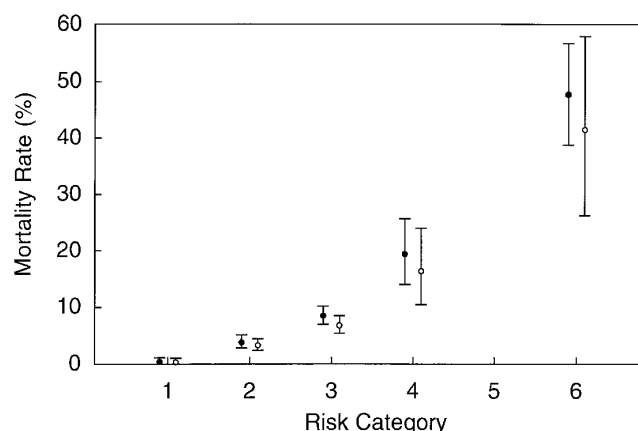
**TABLE 4. Prevalences of additional clinical risk factors**

	PCCC data (n = 4370)	Hospital discharge data (n = 3646)
Age		
≤30 d	833 (19.1%)	450 (12.3%)
31 d–1 y	1383 (31.6%)	1140 (31.3%)
Male	2365 (54.1%)	1926 (52.8%)
Weight ≤2500 g	144 (3.3%)	—
Prematurity	343 (7.8%)	62 (1.7%)
Down syndrome	361 (8.3%)	323 (8.9%)
Presence of other major chromosomal abnormality or recognizable syndrome	184 (4.2%)	104 (2.9%)
Presence of major noncardiac structural anomaly	199 (4.6%)	83 (2.3%)
Transfer from another acute care facility	—	457 (12.5%)

Of these, 3767 (81.9%) underwent a single cardiac procedure; the remaining 835 cases underwent combinations of cardiac procedures. Among the patients with a single procedure, 58 (1.5%) could not be assigned to a risk category. Within the hospital discharge data set, 4493 patients were found. Of these, 3832 (85.3%) underwent a single cardiac procedure and 661 underwent combinations of procedures. Among the patients with a single procedure, 413 (10.8%) could not be assigned to a risk category. Examples of uncategorizable procedures are rare operations that had not been considered by the panel, such as left ventricle to aorta conduit, and procedures with descriptions that were deemed too vague to categorize, such as “revision of procedure” or “other repair of vessel.” With the exception of a small amount of missing weight information in the PCCC data set (0.4%), there were no missing data for the variables analyzed.

For each data set the numbers of cases in each category, along with the in-hospital mortality rates and 95% confidence intervals, are shown in Table 3 and Figure 1. Mortality rates for category 5 are not shown in Figure 1 because of the small number of cases. Note that mortality rates increase monotonically, but not linearly, by risk category. Within each data set, mortality rates for each risk category were different from the rates for all other categories ( $P < .001$  for each data set). Within each risk category, mortality rates did not differ statistically significantly between the two data sets.

The additional clinical factors investigated are shown in Table 4. Age at operation was categorized as at most 30 days, 31 days to 1 year old, and older than 1 year; weight at operation was categorized as less than or equal to 2500 g versus more than 2500 g. Prematurity, Down syndrome,



**Figure 1. Estimated mortality rates (points) and 95% confidence intervals (error bars) by risk category for patients undergoing single cardiac procedure in PCCC (filled circles) and hospital discharge (open circles) data sets.**

presence of a major noncardiac structural anomaly, and presence of a major chromosomal abnormality or syndrome were identified by means of PCCC and *ICD-9-CM* diagnosis codes. Prematurity was defined as gestational age at most 36 weeks in the PCCC data set and as birth weight less than 2500 g or gestational age at most 37 weeks in the hospital discharge data. Major structural anomalies included such conditions as tracheoesophageal fistula and cleft lip or palate. Major chromosomal abnormalities or syndromes included conditions such as DiGeorge syndrome, Turner syndrome, and sickle cell disease. In the PCCC data set the area under the ROC curve for the model containing risk category alone was 0.784. The clinical factor that provided the most additional predictive information about in-hospital mortality was age at operation, followed first by presence of a major structural anomaly and then by prematurity. The area under the ROC curve for the model containing all four of these factors was 0.817. In the hospital discharge data set the area under the ROC curve for the model containing risk category alone was 0.749. Again, the clinical factor that provided the most additional predictive information was age at operation, followed in this case by prematurity. The area under the ROC curve corresponding to the model containing these three factors was 0.815.

In the PCCC data set 835 patients underwent combinations of cardiac procedures. Of these, 174 (20.8%) could not be assigned to a risk category. Among the 661 patients with combinations of procedures in the hospital discharge data set, 434 (65.7%) could not be assigned to a risk category. Patients with combinations of cardiac surgical procedures were initially placed in the category corresponding to the single most complex procedure; the numbers of patients

TABLE 5. Final multivariable models

	PCCC data			Hospital discharge data		
	Odds ratio	95% Confidence interval	P value	Odds ratio	95% Confidence interval	P value
Risk category						
2	6.6	2.4-18.3	<.001	6.8	2.1-22.2	.002
3	15.5	5.7-42.5	<.001	12.7	3.9-40.9	<.001
4	28.2	9.9-80.6	<.001	25.2	7.3-87.3	<.001
5	97.8	10.4-916	<.001	—	—	—
6	93.4	32.4-269	<.001	80.2	20.8-309	<.001
Age						
≤30 d	3.0	2.1-4.1	<.001	4.7	2.9-7.6	<.001
31 d–1 y	1.9	1.4-2.7	<.001	2.5	1.6-3.9	<.001
Major noncardiac structural anomaly	1.8	1.1-2.9	.011	—	—	—
Prematurity	1.8	1.3-2.6	.001	2.9	1.5-6.0	.002
Combination procedures	1.5	1.1-2.0	.009	—	—	—

Model intercepts are −5.76 for the PCCC data and −6.00 for the hospital discharge data.

in each category and the in-hospital mortality rates are shown in Table 3. To explore the issue of how best to adjust for combination procedures, the PCCC and hospital discharge data sets were combined. With combination procedures assigned to the category of the most complex procedure, the area under the ROC curve for the model containing risk category alone was 0.778. This area did not improve, and in fact decreased slightly, when combination procedures were assigned to the highest risk category plus 1. The area also decreased when individual risk categories were summed.

After incorporation of combination procedures by leaving them in the risk category of the most complex procedure, multivariable models including additional clinical factors were reexamined. The same factors entered into the models in the same order in both data sets. In each case the presence of combination procedures was also considered as a binary covariate; this factor added significant predictive information in the PCCC data but not in the hospital discharge data. The final models for each data set are shown in Table 5. The areas under the ROC curves were 0.811 for the PCCC data and 0.814 for the hospital discharge data. According to the Hosmer-Lemeshow test, there were no statistically significant differences between observed and expected numbers of deaths in either data set ( $P = .34$  for PCCC,  $P = .21$  for hospital discharge). The plots of observed versus expected mortality rates are shown in Figure 2.

## Discussion

The task of creating a valid risk-adjustment tool for surgery for congenital heart disease is not a simple one. Because of

the enormous variability inherent in surgery for congenital heart disease, we chose to start with a panel of experts carefully selected for their clinical acumen, research accomplishments, broad experience, and geographic diversity. That panel, after considerable discussion, chose to use surgical procedure as the primary surrogate of risk and therefore divided nearly all surgical procedures for congenital heart disease into six categories of increasing predicted operative risk. They chose to use in-hospital mortality as the outcome of interest because of its overriding clinical importance, its ease of measurement, and its widespread availability in nearly all clinical databases. This initial effort provided considerable power in adjusting for risk of operative mortality, which was further enhanced in two ways. First, the method was applied to two large but different data sets: the high-quality, prospectively collected PCCC data set and the discharge abstract data submitted by hospital record departments to three states (California, Illinois, and Massachusetts). Mortality rates were inspected by the panel of experts, who chose to change the categorizations of a small number of procedures on the basis of the findings. Second, the data sets were also used to empirically identify other risk factors of importance. Additional variables (eg, age, prematurity, major noncardiac structural anomaly) were found to be independently associated with risk and to improve the predictive ability of the risk-adjustment method.

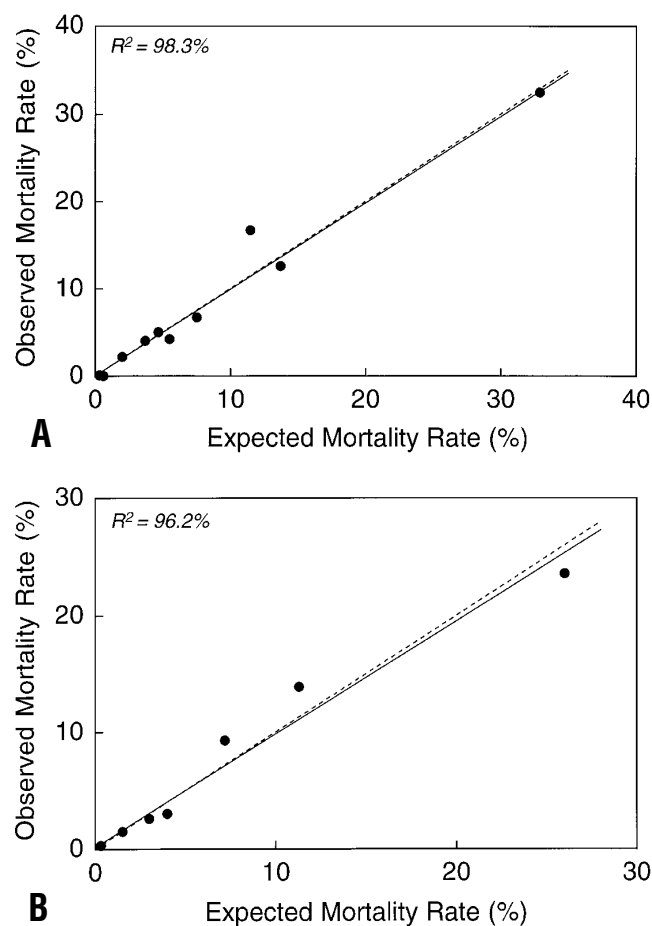
Although many other risk-adjustment techniques have been developed for both adults<sup>10-14</sup> and children,<sup>15-17</sup> few have needed to incorporate the extreme anatomic diversity encountered in procedures for congenital heart disease. Although our method was not specified in advance, our final technique was similar to that applied by the

Department of Veterans Affairs' National Surgical Quality Improvement Program,<sup>18</sup> which also used a score indicating complexity of procedure that was assigned by a panel of experts, and then empirically determined additional risk factors by means of a large, prospectively collected database. Our current methods are a substantial improvement over our previous work<sup>3</sup> because of the input of our panel. These methods differ from those of Hannan,<sup>4</sup> who used a primarily empirical approach that relied on extensive prospective data available through the New York State Cardiac Surgery Reporting System to create four risk categories and then revised the categories according to the advice of an advisory panel.

The overall performance characteristics of our final models compared favorably with adult models attempting to predict mortality for coronary artery bypass graft surgery.<sup>19</sup> The c statistics (areas under ROC curves) for our models were 0.811 with the PCCC data and 0.814 with the hospital discharge data, values somewhat higher than those for models obtained in New York State (0.787),<sup>10</sup> by the Veterans Affairs Continuous Improvement in Cardiac Surgery Study (0.731),<sup>11</sup> and by the Society for Thoracic Surgeons' Coronary Artery Bypass Graft Risk Model (0.800).<sup>12</sup> Our results were only slightly lower than those obtained by Hannan for surgery for congenital heart disease in New York (0.818),<sup>4</sup> despite the fact that this study and the others mentioned used extensive amounts of prospectively collected data. Model calibration was also equivalent to or better than that for these other models.<sup>4,10,12</sup>

One strength of our methodology relative to a purely empirical approach is shown by our ability to incorporate relatively rare procedures. This is especially important because pediatric cardiac surgery is characterized by extreme diversity. The data sets that we used for this project are among the largest available for this type of work, yet few procedures would have been placed uniquely into one category on the basis of estimated mortality rates and confidence intervals. By using judgment-based methods, we were able to group procedures with extremely few cases, such as repair of partially anomalous pulmonary venous connection or atrial switch operation, into risk categories. This allowed a higher proportion of the total caseload to be included.

At present, despite its limitations, our method should prove useful in helping program directors and other similarly interested parties to gain an enhanced understanding of the overall outcomes that they are achieving. Information regarding overall performance is difficult to obtain because of the extreme diversity of conditions that comprise congenital heart disease. Unlike in adult heart surgery, even the most common lesions make up only a small fraction of most surgical caseloads. Procedure-specific outcome information is thus intrinsically limited in two regards. First the precision of measurement is limited by small sample sizes.



**Figure 2. Observed versus expected mortality rates within groups of patients with similar predicted probabilities of death constructed for PCCC (A) and hospital discharge (B) data sets. Solid lines were fitted with weighted linear regression; dashed lines represent ideal case in which observed and expected rates are equal.**

Second, performance measures for individual defects may not be reflective of overall performance. Measurement tools that include all or most of a total surgical caseload should provide a more precise and better reflection of overall performance. With this methodology, even for a relatively poor source of data such as hospital discharge abstract data,<sup>20-22</sup> approximately 80% of a total caseload can be included for the evaluation of relative performance. For richer sources of data, such as the prospectively collected information from the PCCC, almost 98% of the caseload can be included. Once validated, RACHS-1 can also be used in the comparative assessment of outcomes between institutions to guide quality improvement efforts.

As with all risk-adjustment tools, one must carefully distinguish appropriate and inappropriate uses. From the onset,



our method was not created to predict the risk of death for individual patients, but rather to be a tool that would allow meaningful comparisons across groups of patients. Many relevant but unusual clinical factors, such as concomitant preoperative pneumonia, were not included by the panel. Although such factors might be important determinants of outcome for individuals, they were deemed to be an unlikely source of confounding among groups of patients, either because of their infrequency or because they would be expected to be similarly represented within patient groups. Furthermore, the RACHS-1 method was derived and validated to provide risk adjustment for comparisons of in-hospital mortality. No assumptions can be made about the method's performance for other outcomes, such as morbidity or late mortality. Also, we were careful to avoid including factors such as insurance type that might be associated with mortality but are not a part of the intrinsic clinical risk for death. By using our method to adjust for intrinsic risk differences only, researchers will be better able to make inference about differences related to type of insurance or to other factors that might reflect important information about our health care system. Last, and most importantly, our methodology allowed the panel to refine RACHS-1 on the basis of the observed mortalities in the PCCC and hospital discharge data sets. Thus the performance characteristics we obtained may be overly optimistic. The true performance of RACHS-1 will not be known until the method has been tested in an independent data set. This work is currently in progress.

In summary, our RACHS-1 method has several useful features. First, it can be applied to nearly all cases of surgery for congenital heart disease in childhood. Second, it can be used both with high-quality, prospectively collected data sets and with lower quality hospital discharge data sets. It relies on variables (in-hospital mortality, surgical procedure codes, age, prematurity, major noncardiac structural anomaly) that are easy to identify, quantify, and collect. The predictive ability of RACHS-1 compares well with those of other risk-adjustment methods. Finally, it is a tool that can easily adapt to changes in the field of pediatric cardiac surgery. By reconvening the panel (or a similar one), new procedures can be categorized, old procedures can be recategorized, and revised categories can be tested on newer data sets and refined further.

We are indebted to the panel of experts and thank them for their participation.

## References

1. Erickson LC, Wise PH, Cook EF, Beiser A, Newburger JW. The impact of managed care insurance on use of lower-mortality hospitals by children undergoing cardiac surgery in California. *Pediatrics*. 2000;105:1271-8.
2. Chang RK, Chen AY, Klitzner TS. Factors associated with age at operation for children with congenital heart disease. *Pediatrics*. 2000;105:1073-81.
3. Jenkins KJ, Newburger JW, Lock JE, Davis RB, Coffman GA, Iezzoni LI. In-hospital mortality for surgical repair of congenital heart defects: preliminary observations of variation by hospital caseload. *Pediatrics*. 1995;95:323-30.
4. Hannan EL, Racz M, Kavey RE, Quaegebeur JM, Williams R. Pediatric cardiac surgery: the effect of hospital and surgeon volume on in-hospital mortality. *Pediatrics*. 1998;101:963-9.
5. Sollano JA, Gelijns AC, Moskowitz AJ, Heitjan DF, Cullinane S, Saha T, et al. Volume-outcome relationships in cardiovascular operations: New York State, 1990-1995. *J Thorac Cardiovasc Surg*. 1999;117:419-30.
6. Clancy RR, McGaurn SA, Wernovsky G, Spray TL, Norwood WI, Jacobs ML, et al. Preoperative risk-of-death prediction model in heart surgery with deep hypothermic circulatory arrest in the neonate. *J Thorac Cardiovasc Surg*. 2000;119:347-57.
7. Gutgesell HP, Massaro TA, Kron IL. The arterial switch operation for transposition of the great arteries in a consortium of university hospitals. *Am J Cardiol*. 1994;74:959-60.
8. Daley J, Schwartz M. Developing risk-adjustment methods. In: Iezzoni LI, editor. Risk adjustment for measuring healthcare outcomes. 2nd ed. Chicago: Health Administration Press; 1997.
9. Moller JH. Perspectives in pediatric cardiology, vol 6. Surgery of congenital heart disease: Pediatric Cardiac Care Consortium 1984-1995. New York: Futura Publishing; 1998. p. 279-329.
10. Hannan EL, Kilburn H, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York state. *JAMA*. 1994;271:761-6.
11. Grover FL, Johnson RR, Marshall G, Hammermeister KE. Factors predictive of operative mortality among coronary artery bypass subsets. *Ann Thorac Surg*. 1993;56:1296-307.
12. Shroyer AL, Grover FL, Edwards FH. 1995 coronary artery bypass risk model: The Society of Thoracic Surgeons Adult Cardiac National Database. *Ann Thorac Surg*. 1998;65:879-84.
13. Moscucci M, O'Connor GT, Ellis SG, Malenka DJ, Sievers J, Bates ER, et al. Validation of risk adjustment models for in-hospital percutaneous transluminal coronary angioplasty mortality on an independent data set. *J Am Coll Cardiol*. 1999;34:692-7.
14. Polanczyk CA, Rohde LE, Philbin EA, Di Salvo TG. A new casemix adjustment index for hospital mortality among patients with congestive heart failure. *Med Care*. 1998;26:1489-99.
15. Pollack MM, Koch MA, Bartel DA, Rapoport I, Dhanireddy R, El Mohandes AA, et al. Comparison of neonatal mortality risk prediction models in very low birth weight infants. *Pediatrics*. 2000;105:1051-7.
16. Richardson DK, Gray JE, McCormick MC, Workman K, Goldmann DA. Score for neonatal acute physiology: a physiologic severity index for neonatal intensive care. *Pediatrics*. 1993;91:617-23.
17. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated pediatric risk of mortality score. *Crit Care Med*. 1996;24:743-52.
18. Khuri SF, Daley J, Henderson W, Hur K, Demakis J, Aust JB, et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. *Ann Surg*. 1998;228:491-507.
19. Iezzoni LI, Ash AS, Schwartz M, Daley J, Hughes JS, Mackiernan YD. Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. *Am J Public Health*. 1996;86:1379-87.
20. Romano PS. Can administrative data be used to compare the quality of health care? *Med Care Rev*. 1993;50:451-77.
21. Hannan EL, Kilburn Jr H, Lindsey ML, Lewis R. Clinical versus administrative data bases for CABG surgery: does it matter? *Med Care*. 1992;30:892-907.
22. Romano PS, Roos LL, Luft HS, Jollis JG, Doliszny K. A comparison of administrative versus clinical data: coronary artery bypass surgery as an example. Ischemic Heart Disease Patient Outcomes Research Team. *J Clin Epidemiol*. 1994;47:249-60.